Enhancing Accuracy in Predicting Thailand's Rice Exports: A Hybrid Modeling Approach

Paothai Vonglao, Kajita Somnat, Somporn Thepchim and Thanakon Sutthison*

Program of Applied Statistics, Faculty of Science, Ubon Ratchathani Rajabhat University, Ubon Ratchathani Province, 34000, Thailand

* Corresponding author. E-mail address: thanakon.s@ubru.ac.th

Received: 19 July 2022; Revised: 3 August 2023; Accepted: 9 August 2023; Available online: 16 November 2023

Abstract

Thailand's rice exports are currently experiencing a declining trend in relation to the proportion of rice production. This calls for the need to accurately predict future developments, which holds immense importance for stakeholders involved. Accurate predictions enable the formulation of effective policies and strategies to boost Thailand's rice exports in the future. To address this objective, this research aims to identify a suitable model for forecasting the monthly quantity of Thailand's rice exports. A sophisticated hybrid model is proposed, integrating the strengths of Empirical Mode Decomposition (EMD), Seasonal Auto-Regressive Integrated Moving Average (SARIMA), and Support Vector Regression (SVR). The model's parameters are optimized using Genetic Algorithm (GA) to ensure optimal performance. To evaluate the hybrid model's effectiveness, rigorous performance criteria are employed, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide a comprehensive assessment of the model's predictive capabilities and overall performance. The research findings demonstrate that the developed hybrid model outperforms individual models across all performance criteria. This solidifies its reliability in generating accurate forecasts for Thailand's monthly rice export quantities. Consequently, the hybrid model emerges as a valuable tool for organizations seeking to proactively forecast and effectively manage the dynamics of Thailand's rice exports in the future.

Keywords: Hybrid model, Empirical mode decomposition, Support vector regression, Genetic algorithm, SARIMA model

Introduction

Rice holds a significant position as a staple food consumed worldwide, and Thailand proudly stands as one of the leading exporters, accounting for one-fourth of the world's major rice exports. The expanse of rice cultivation in Thailand spans an impressive 61 61 875 000 rai or 9,900,000 hectares. Analyzing the historical data of rice production over the past five years (from 2016 to 2020), we observe varying figures ranging from 19.20 to 20.577 million tons. Within the country, rice consumption remains stable, ranging from 11.000 to 12.700 million tons, while the trend in rice exports tells a different story, showing a gradual decline from 11.615 to 5.706 million tons (Office of Agricultural Economics, 2022). The intricate interplay between rice production, exports, and the dynamics of the global rice market can be fascinating to explore. Examining the available data from 2016 to 2021, we note subtle fluctuations in Thai rice production, likely influenced by external factors such as ever-changing weather patterns, severe droughts, and destructive floods (Jeong, Ko, & Yeom, 2022). While domestic rice consumption in Thailand shows a modest upward trajectory, the consistent decline in rice exports poses an intriguing topic worthy of investigation. A reliable and accurate forecast of rice exports can prove immensely valuable, aiding relevant organizations in formulating strategies to ensure timely and sustainable rice imports or exports. Numerous methods exist for predicting Thai rice exports, with the prevailing approach relying on statistical forecasting. These techniques leverage time series data to develop models that offer insights into future trends, facilitating effective planning and decision-making. Notably, studies highlight the use of models such as Box-Jenkins, Winters, additive exponential smoothing, and combined models for forecasting jasmine rice exports of Thailand (Keerativibool, 2014). Furthermore, Co and Boosarawongse (2007) have employed advanced artificial neural network models to forecast Thai rice exports, considering the intricate and non-linear nature of export data. However, it is worth noting that creating accurate forecasting models for rice exports poses significant challenges (Huang, Hasan, Deng, & Bao, 2021). Conventional statistical methods like Auto-Regressive Integrated Moving Average (ARIMA) or Seasonal Auto-Regressive Integrated Moving Average (SARIMA) may fall short of capturing the intricate dynamics of time series data, resulting in lower forecast accuracy. Consequently, non-linear forecasting models such as Artificial Neural Networks (ANN) and Support Vector Regression (SVR) have gained prominence, showcasing higher accuracy and efficiency compared to traditional approaches (Basir Chowdhury, Islam, & Ashik-E-Rabbani, 2021; Chen, Chen, & Jiang, 2021; Fan, Yu, Dong, Yeh, & Hong, 2021; Feng et al., 2020; Gu et al., 2016; Liu, Ma, Wang, Lu, & Lin, 2021; Su, Xu, & Yan, 2017; Tang, Bai, Yang, & Lu, 2020; Yang, Che, Deng, & Li, 2019). However, it is essential to recognize that no single forecasting model can universally predict time series data with absolute precision. Hence, researchers propose hybrid models for time series forecasting, blending the strengths of both linear and non-linear models to enhance the precision of forecasts. Consequently, hybrid models have gained popularity and widespread usage in current forecasting practices (Abdollahi, 2020; Prado, Minutolo, & Kristjanpoller, 2020; Zhang, Ding, & Sun, 2020).

In situations where time series data exhibit significant variability or are influenced by various factors, the precision of forecasting models can be affected. Combining linear and non-linear models to address these factors and achieve optimal parameters poses a challenge due to the risk of overfitting. To overcome this obstacle, numerous researchers have turned to the Empirical Mode Decomposition (EMD) method, originally proposed by Huang (1971) and subsequently utilized by Duan, Han, Huang, Zhao, and Wang (2016), Meng et al. (2019), Yaslan and Bican (2017). EMD offers a solution by effectively decomposing time series data into distinct components, thereby mitigating noise and enhancing the efficiency of forecasting (Kao, Nawata, & Huang, 2020). Within the realm of time series analysis, noise elements such as trends, seasonality, and unknown factors can significantly impact the performance of forecasts. Consequently, to elevate the accuracy of predictions, forecasters must diligently address the challenges associated with managing noise in time series data. There exist various approaches to tackle these issues and improve forecasting models. One such approach involves fine-tuning hyperparameters for algorithms like the Support Vector Regression (SVR) model, which has demonstrated efficacy in enhancing prediction accuracy (Kao et al., 2020). Another viable strategy entails decomposing the time series data into components using EMD, effectively reducing noise, and subsequently utilizing the segregated data to generate more precise forecasts using time series models (Kao et al., 2020; Fan et al., 2022; Meng et al., 2019). Employing this methodology effectively addresses the complexities associated with intricate time series data.

Considering this discussion, it is evident that a hybrid model combining EMD and other techniques can effectively address the challenges posed by volatile and complex time series data, resulting in improved forecasting accuracy. Therefore, the research team proposed a new hybrid model that integrates EMD, SARIMA, Genetic Algorithm (GA), and SVR to forecast Thai rice exports. This model provides an alternative solution for relevant organizations to forecast rice exports in the future. The hybrid model involves decomposing Thai rice export time series data into intrinsic mode functions (IMFs) and a residual term using EMD. IMFs represent the oscillatory components of the time series data, and the IMFs, along with the residual term, are used to build the SARIMA model. However, when the time series data exhibit non-linear characteristics, the SARIMA model's forecasting accuracy is compromised. To address this limitation, the SVR model, known for its effectiveness in predicting non-linear time series data, is utilized (Kao et al., 2020). Nevertheless, the SVR model's forecasting accuracy heavily depends on setting suitable parameters. Therefore, the Genetic Algorithm (GA) is employed to find the optimal parameters for the SVR model (Xu & Zhang, 2021; Kao et al., 2020). This research study combines EMD, SARIMA, GA, and SVR to create a hybrid model that forecasts the IMFs of Thai rice export data, and the forecasts from each component are aggregated in the final step.

The study focuses on three sets of Thai rice export data: jasmine rice, Pathum Thani rice, and sticky rice. These rice varieties are widely grown and exported in large quantities by Thailand. The models used in this research study include SARIMA, GA-SVR, EMD-SARIMA, EMD-GA-SVR, and EMD-SARIMA-GA-SVR. The selection of the appropriate model is based on evaluation criteria such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). By using these criteria, the differences in forecasting accuracy among the proposed models can be assessed. The model with the lowest MAE, RMSE, and MAPE values indicates higher accuracy and suitability for future rice export predictions in Thailand.

Methods and Materials

1. Data used in the experiment

Three time series datasets of Thailand's rice exports, namely jasmine rice, Pathum Thani rice, and sticky rice (Unit: kg) from January 2011 to March 2022 published on the website of the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives, a total of 135 values, were used in the experiment (Office of Agricultural Economics, 2022) as presented in Figure 1. The data were divided into 2 sets. The first set was the training dataset, which was the data from January 2011 to December 2019, with 108 values. This dataset was used to create the model. The second set was the testing dataset, which was the data from January 2020 to March 2022, with 27 values. This dataset was used for model testing. The criteria for dividing data to measure forecast performance (cross-validation) were based on 80% and 20% according to the concept of Gholamy Kreinovich, and Kosheleva (2018)

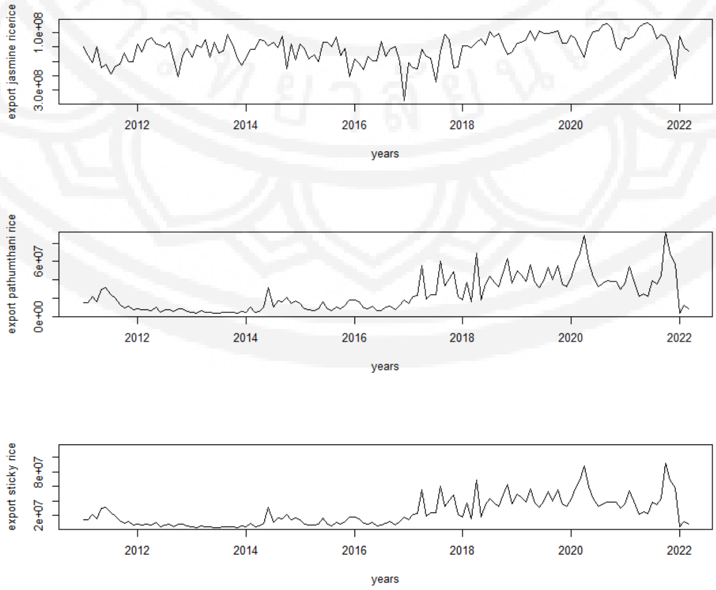

Figure 1 Original three time series datasets of Thailand's rice exports from January 2011 to March 2022

2. Seasonal autoregressive integrated moving average (SARIMA)

SARIMA is the most popular linear model for forecasting seasonal time series, presented by Box Jenkins. The forecast equation is as follows (Box, Jenkins, Reinsel, & Ljung, 2008):

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{1}$$

$$\delta = \mu\phi_p(B)\Theta_P(B^s) \tag{2}$$

$$\phi_p(B) = 1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p \tag{3}$$

$$\Phi_P(B^s) = 1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps} \tag{4}$$

$$\theta_q(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q \tag{5}$$

$$\Theta_Q(B^s) = 1-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_Q B^{Qs} \tag{6}$$

Y_t is the observed value at a time t, ε_t is the noise component assumed to be NID$(0,\sigma^2)$, and t is the time ranging from 1 to n. s is the time unit in 1 season, d and D are the orders of differences and seasonal differences. B is a backward operator. $B^s Y_t = Y_{t-s}$ and $BY_t = Y_{t-1}$, $\phi_p(B)$ is non-seasonal autoregressive operator of order p. $\Phi_P(B^s)$ is the seasonal autoregressive operator of order P. $\theta_q(B)$ is non-seasonal moving average operator of order q. $\Theta_Q(B^s)$ is the seasonal moving average operator of order Q.

To create a SARIMA model, SARIMA (p,d,q)(P,D,Q)s must be defined. The researchers wrote the commands using an R programming language with the auto.arima() function in the Package "forecast" (R Core Team, 2021) to define the suitable ARIMA(p,d,q)(P,D,Q)s. The model construction process consisted of 4 following steps.

Step 1: The time series consistency must be checked by considering the graphs of that time series. If the time series graphs have a tend component, that is, the data increase or decrease proportionally to time and is unevenly distributed, the time series data are inconsistent. In this study, the data inconsistency was tested by Augmented Dickey-Fuller (ADF). If the test result is not statistically significant, it means that the time series data are inconsistent. Therefore, it must be converted to consistent data by finding out the logarithm of the data and differentiating it to stabilize the time series data before it is used to define the model.

Step 2: The forecasting model was defined.

Step 3: The parameters were estimated, and the suitability of the model was verified. When obtaining a suitable model, before applying the model to forecasting, it must be checked with 4 diagnostic checking lists. (1) Zero mean error was tested by the t-test. (2) The normal distribution error was tested by the Kolmogorov-Smirnov. (3) The interdependence of the error was tested by Ljung-Box Q-statistics. (4) The error must have constant variance determined by plotting the graph to consider the distribution between the error and the forecast value. If the graph is distributed with no pattern, the error has constant variance.

Step 4: The suitable model was used to forecast Thailand's monthly rice exports.

3. Empirical mode decomposition (EMD)

EMD is an empirical signal processing algorithm used for extracting features of nonlinear data. Since EMD does not need to set the parameters, it can automatically fit into the rice export data. To reduce the parameters of the proposed model and improve the robustness of the proposed model, in this paper, EMD is selected to decompose the export rice data into several sub-layers. The computational steps of EMD can be described as follows (Duan et al., 2016; Qiu, Ren, Suganthan, & Amaratunga, 2017; Meng et al., 2019; Mi, Liu, & Li, 2019; Yaslan & Bican, 2017):

(1) Determine all the local extremes of the rice export data $X(t)$.

(2) Connect all the local maxima to calculate the upper envelope $X_U(t)$ by using a cubic spline line. Similarly, connect all the local minima to calculate the lower envelope $X_L(t)$.

(3) Compute the mean envelope $M(t)$:

$$M(t) = [X_U(t) + X_L(t)]/2 \tag{7}$$

(4) Calculate the variable $Y(t)$:

$$Y(t) = X(t) - M(t) \tag{8}$$

If $Y(t)$ is the IMF (Intrinsic Mode Function), set $c(t) = Y(t)$; otherwise, replace $X(t)$ by $Y(t)$, and repeat steps (1) - (4), until the termination condition is satisfied.

(5) Calculate the residual $R(t)$:

$$R(t) = X(t) - C(t) \tag{9}$$

replace $X(t)$ with $R(t)$, and repeat steps (1) - (5), until all the IMF are found.

4. Support vector regression (SVR)

At present, the SVR method is one of the most popular and high-accuracy methods (Chen et al., 2021; Gu et al., 2016; Fan et al., 2021; Feng et al., 2020; Su et al., 2017; Tang et al., 2020), introduced by Vapnik (1998). It minimizes structural risks by adjusting the low-dimension dataset on the input space to be in the high-dimension dataset on the feature space using a kernel function as presented in the forecast Equation (10)

$$f(x) = w^T \phi(x) + b \tag{10}$$

When w is the vector weight, and b is the error of the regression line, the values of w and b are determined by investigating the minimum value from Equation (11).

$$R(\mathrm{C}) = \frac{1}{2}\|w\|^2 + C.\frac{1}{n}\sum_{i=1}^{N}\left|y_i - f(x)\right|_{\varepsilon} \tag{11}$$

Using the SVR method to predict the output value from the input vector, an epsilon tube is constructed using various types of loss functions, with the most popular being ε-insensitive introduced by Vapnik (1998) because it is a common loss function used in time series forecasting (Chen et al., 2021; Fan et al., 2021; Feng et al., 2020; Tang et al., 2020). The ε-insensitive loss function is shown in Equation (12).

$$\left|y_i - f(x)\right| = \begin{cases} 0 & if \left|y_i - f(x)\right| \le \varepsilon \\ \left|y_i - f(x)\right|_{\varepsilon} & otherwise \end{cases} \tag{12}$$

ξ_i and ξ_i^* are the slack variables and are away from the edge of epsilon (ε-tube) as shown in Figure 2.

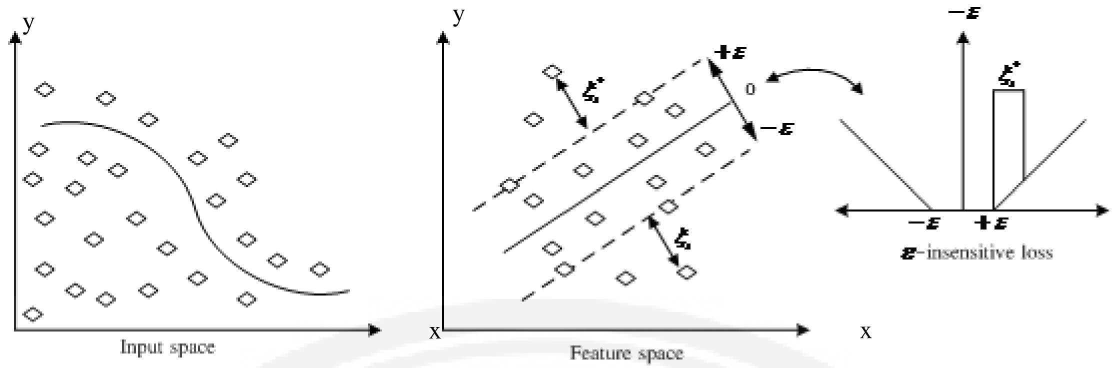

Figure 2 Boundary of Support Vector Regression

Adding ξ_i and ξ_i^* is the measurement of all deviations of the training dataset outside the edge of the plane or the error of data from the upper and lower edges of the plane. Therefore, adding ξ_i and ξ_i^* causes Equation (11) to be converted to the constrained condition as in Equation (13).

$$\text{Minimize} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{13}$$

$$\text{Subject to} \begin{cases} y_i - w\phi_i - b \le \xi_i + \varepsilon \\ -y_i + w\phi_i + b \le \xi_i^* + \varepsilon \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{14}$$

Solving Equation 13 with the constrained condition of Equation 14 can be done by converting it to the dual problem with Lagrange multipliers as shown in Equation 15 and the condition as in Equation 16.

$$\mathit{Maximize}\ (\alpha_i, \alpha_i^*) = \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \tag{15}$$

$$\mathit{Subject\ to} \begin{cases} \sum_{i=1}^{N}(\alpha_i + \alpha_i^*) = 0 \\ 0 \le \alpha_i \le C, \quad i = 1,\ldots,N \\ 0 \le \alpha_i^* \le C,\ i = 1,\ldots,N \end{cases} \tag{16}$$

When x_i, x_j are the input data, α_i, α_i^* are Lagrange multipliers, C is constant. N is the number of support vectors. If the input is a support vector, $\alpha_i, \alpha_i^* > 0$. If the input vector is not a support vector, $\alpha_i, \alpha_i^* = 0$. After calculating α_i and α_i^* from the learning dataset, the SVR equation can be created to predict the output from the input vector as shown in Equation 17.

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{17}$$

When α_i and α_i^* are Lagrange multipliers, $k(x_i, x)$ is a kernel function used to convert data to a higher dimension. There are a variety of kernel functions that are commonly used in support vector regression. This research applied the Gaussian (RBF) kernel: $k(x_i, x) = \exp(-\frac{\|x_i . x\|^2}{2\sigma^2})$ since it was the most accurate.

5. Genetic algorithm (GA)

GA is part of the metaheuristic method used to discover suitable parameters for a forecasting model, in particular, forecasting by machine learning methods (Prado et al., 2020; Kao et al., 2020), to increase the accuracy of forecasting values. GA is a concept proposed by Holland (1975). Optimal answer determinations begin with a solution randomly assigned to the initial population. Each answer is encoded and called a chromosome. Each chromosome contains a gene, which is the parameter that needs to be optimized. Then the answer from the previous generation of the population, known as the parent, will be used to create the next generation, known as the offspring. Genetic operators can be divided into two types: crossover and mutation. In the processing, the offspring will then replace the parent and this will be repeated until the stop condition is reached. To create forecast equations using SVR, three important parameters affect forecast accuracy: parameter C, parameter σ and parameter ε. The process of investigating suitable parameters of the SVR method by GA can be carried out as follows.

5.1 Chromosome representation: This is the process of identifying the gene under chromosomes. When it is used to find the appropriate parameters for the SVR method, each chromosome contains the gene representing the parameter. The chromosome is encoded as the floats chromosome. Then a random initial population is generated for further fitness evaluation.

5.2 Fitness evaluation: This is the evaluation of the suitability of each chromosome for the selection of offspring chromosomes to be the next generation of the population.

5.3 Selection operation: This is the step in selecting the offspring chromosomes to be the next generation of the population.

5.4 Crossover: This is the creation of offspring chromosomes by using two parent chromosomes to perform crossover at a specified rate of crossover.

5.5 Mutation: It is the creation of offspring chromosomes by using one parent chromosome to perform mutation.

5.6 Replacement: It is to replace the original population with a new, more suitable population. Steps 5.2 to 5.6 are repeated until the GA stop condition is reached.

6. Hybrid model

The combination of the SARIMA and SVR models represents a powerful and efficient approach for forecasting time series data. Each model exhibits distinct strengths. SARIMA excels in accurately predicting linear time series data while facing limitations when it comes to non-linear and non-stationary data. To address these challenges, the EMD technique is applied, enabling the transformation of non-linear data into a stationary form. Subsequently, SARIMA is utilized to forecast the decomposed components obtained through EMD. Although EMD effectively handles non-linearity and reduces complexities in the time series data, there are often hidden factors that impact its accuracy. To enhance the performance of the EMD-SARIMA model, this research integrates SVR to forecast the residuals and incorporates them into a comprehensive hybrid model called EMD-SARIMA-SVR. As the SVR model necessitates appropriate parameter tuning, the GA method is employed to identify optimal parameter settings, ultimately refining the effectiveness of the hybrid model. The overall process of model development can be summarized as follows.

6.1 EMD-SARIMA model

To develop the EMD-SARIMA hybrid model, the researchers analyzed the data on rice exports using the EMD method (details presented in Section 3). The objective was to decompose the non-linear time series data into a stationary form, known as the Intrinsic Mode Function (IMFs) and residual. Each separated component is then divided into two sets: a training dataset of 108 values and a testing dataset of 27 values (details provided in Section 1). The training dataset is utilized for forecasting using the SARIMA model (details in Section 2). Once a suitable model is obtained, it is used to forecast 27 future values. These predicted values from all components are then combined in the final step. Subsequently, performance metrics are calculated and sequentially presented in Table 5.

6.2 EMD-GA-SVR model

To develop the EMD-GA-SVR hybrid model, the researchers analyzed the data on rice exports using the EMD method, similar to the EMD-SARIMA model. The objective was to decompose the non-linear time series data into a stationary form. Subsequently, the researchers obtained the IMF (Intrinsic Mode Functions) and residuals. These components were divided into two sets: a training dataset of 108 values and a testing dataset of 27 values (details provided in Section 1). The training dataset was utilized to make predictions using the SVR model (details in Section 4). The following steps can be followed:

6.2.1 The training dataset was transformed into a dataset $D=\{(x_t, y_t)\}_{i=1}^{n}$ by rearranging the time series data into a lagged matrix form with m columns as follows (Sujjaviriyasup, 2021).

$$y_1, y_2, y_3, \ldots, y_{t-3}, y_{t-2}, y_{t-1}, y_t$$

$$\begin{bmatrix} y_1, & y_2, & y_3, & \cdots & y_{m-2}, & y_{m-1}, & y_m \\ y_2, & y_3, & y_4, & \cdots & y_{m-1}, & y_m, & y_{m+1} \\ y_3, & y_4, & y_5, & \cdots & y_m, & y_{m+1}, & y_{m+2} \\ \vdots & & & \vdots & & & \vdots \\ y_{t-m+1}, & y_{t-m+2}, & y_{t-m+3}, & \cdots & y_{t-2}, & y_{t-1}, & y_t \end{bmatrix}$$

From the metrics, the data are displayed backwards from column 1 to column $m-1$, used as input data, and the column m is the target data. This is done to enable the SVR model to learn and create a predictive function for the characteristics of the data. This research selects several columns ranging from 2 to 12, as the researchers collected monthly data. There are 12 periods to reflect the calendar cycle. If a large volume of backward data is used, the forecasted values may not be appropriate. The SVR model is implemented by programming in the R language, which has the SVM() function in the e1072 package presented by Meyer et al. (2015), and the GA method, which is implemented by the GA package published by Scrucca (2013), is used to search for the most suitable parameters for the SVR model. The process of finding the appropriate parameter values is described in detail (as shown in Figure 3).

(1) Initial Population is defined by randomly selecting chromosomes to be members of the population, with a total of 100 chromosomes. Each chromosome consists of 3 genes, where each gene represents the parameter values of the SVR model (C, σ^2, and ε).

(2) Randomize parameters for experimental use with data to obtain the most appropriate values. This research is divided into 5 parts (5-fold cross-validation on the training dataset) and each part randomly tests the data with parameters to obtain the most suitable parameters for the data.

(3) Assess the fitness value of chromosomes to select offspring to form the next generation population. The evaluation function is defined using the RMSE (Root Mean Square Error) as a measure of suitability.

(4) Check the termination condition of the GA method. This research project specifies the termination condition when the population has reached 100 generations.

(5) If the termination condition of the GA method's execution has not been met, proceed with the selection process, which is a step to select offspring chromosomes to become the population in the next generation.

(6) Perform Crossover, which is the process of creating offspring chromosomes by crossing over the parental chromosomes, specifically using 3 chromosomes. The Crossover probability is set to 0.80.

(7) Perform mutations, which involve generating offspring chromosomes by taking one chromosome from both parents. The mutation rate is set to 0.1.

(8) Replacement is the process of replacing the original population with a new population that is better suited. It involves replacing the parental chromosomes with offspring chromosomes according to a crossbreeding rate of 0.80.

(9) The search boundaries for C, σ^2 and ε are within the intervals $[10^1, 10^2]$, $[10^{-4}, 10^{-3}]$, and $[10^{-5}, 10^{-4}]$ respectively.

(10)Generate the next generation of the SVR model using the GA method and repeat steps 2.2 - 2.9 until reaching the termination condition of the GA method.

6.2.2 Obtain the optimal parameter values of the SVR method from the GA method, then construct the best SVR model using the training dataset.

6.2.3 Predict the components of IMFs and the data residuals for each set using the input vectors of the testing dataset, consisting of 27 values, from the SVR model obtained in step 3.

6.2.4 Combine the forecast values of each component to obtain the forecast data of Thailand's rice exports for all three datasets. Then, proceed to compare the forecasts in the next order. The details are shown in Table 5

6.3 EMD-SARIMA-GA-SVR model

To construct the EMD-SARIMA-GA-SVR model, the researchers synergistically combined the capabilities of the SARIMA and SVR models for accurate forecasting. The model creation process entails analyzing the time series data of rice exports using the EMD method, mirroring the methodologies employed in the EMD-SARIMA and EMD-GA-SVR models, to transform the inherent non-linear characteristics into a stationary format. The acquired Intrinsic Mode Functions (IMFs) and residual values are subsequently partitioned into two distinct sets: a training dataset of 108 values and a testing dataset of 27 values. Subsequently, the training dataset is employed to generate forecasts using the SARIMA-GA-SVR model, as illustrated in Figure 4. The detailed steps of the model creation process are as follows:

6.3.1 Analyze the data using the Empirical Mode Decomposition (EMD) method to separate the components of the original time series into a set containing the Intrinsic Mode Functions (IMFs) and residual values.

6.3.2 Divide the obtained IMFs and residual values into two segregated sets: a training dataset of 108 values and a testing dataset of 27 values.

6.3.3 Leverage the SARIMA model, expounded upon in Section 2, to forecast the training dataset, thereby capturing the linear characteristics of the data.

6.3.4 Calculate the residuals by deducting the actual values from the forecasts, facilitating the creation of a non-linear relationship function. This function accurately elucidates the non-linear relationship characteristics of the residuals obtained from the well-suited SARIMA model.

6.3.5 Utilize the residuals derived from step 6.3.4 to construct the GA-SVR model, meticulously explicated in Section 6.2, and effectively forecast 27 future values.

6.3.6 Merge the forecasts derived from steps 6.3.3 and 6.3.5, subsequently computing the performance metrics. Comprehensive details about these metrics are sequentially presented in Table 5 and Table 6.

7. The benchmarks for evaluating the performance of the model

The performance of the model was evaluated by the following 3 benchmarks (Chen et al., 2021):

7.1 mean absolute error (MAE)

$$MAE = \sum_{t=1}^{n} \left| Y_t - \hat{Y}_t \right| / n \tag{18}$$

7.2 root mean square error (RMSE)

$$RMSE = \sqrt{\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2 / n} \tag{19}$$

7.3 mean absolute percentage error (MAPE)

$$MAPE = 100 \times \sum_{t=1}^{n} \left| 1 - \hat{Y}_t / Y_t \right| / n \tag{20}$$

If MAE, RMSE, and MAPE values are low, the model is highly efficient.

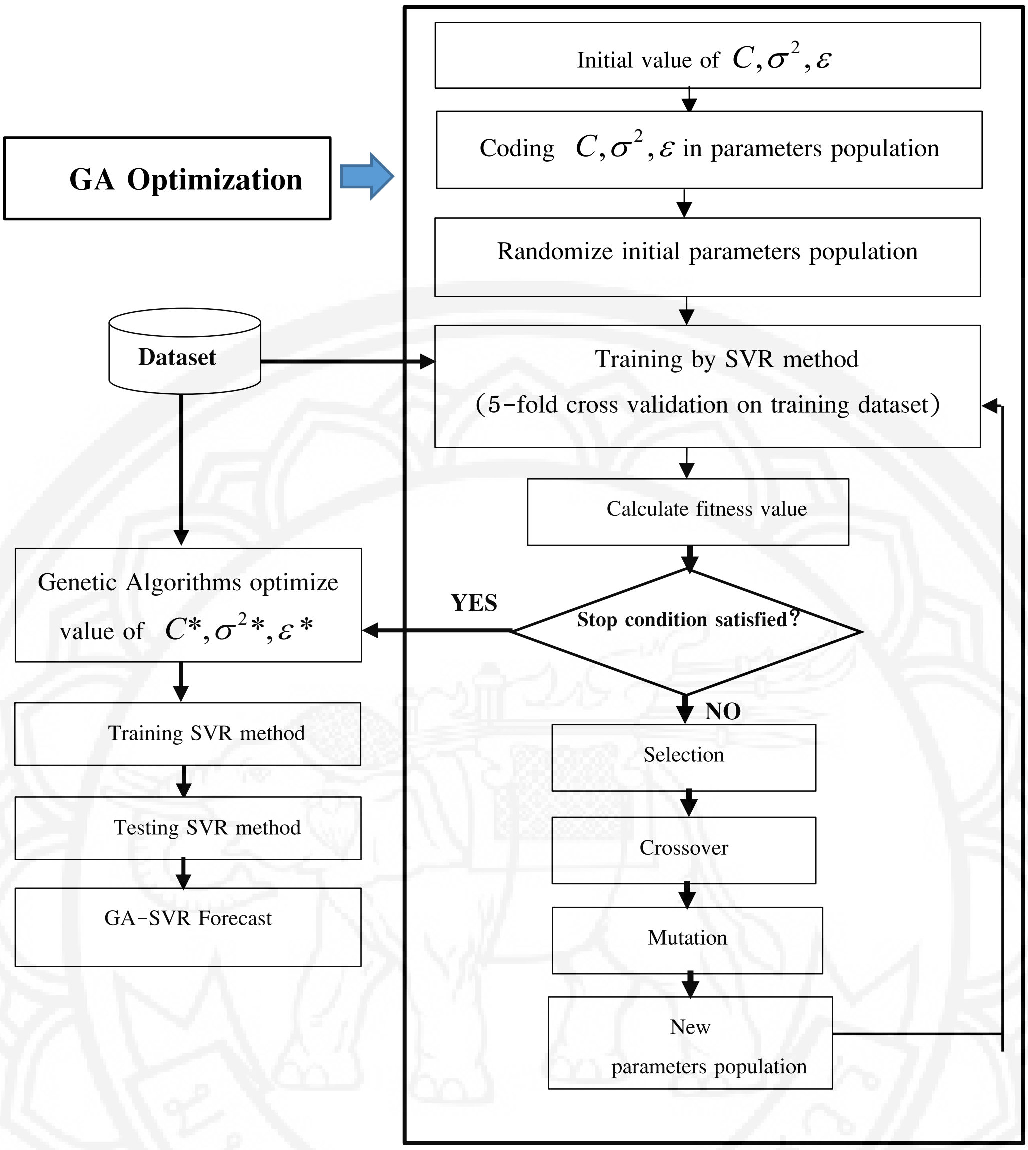

Figure 3 GA-SVR Optimization

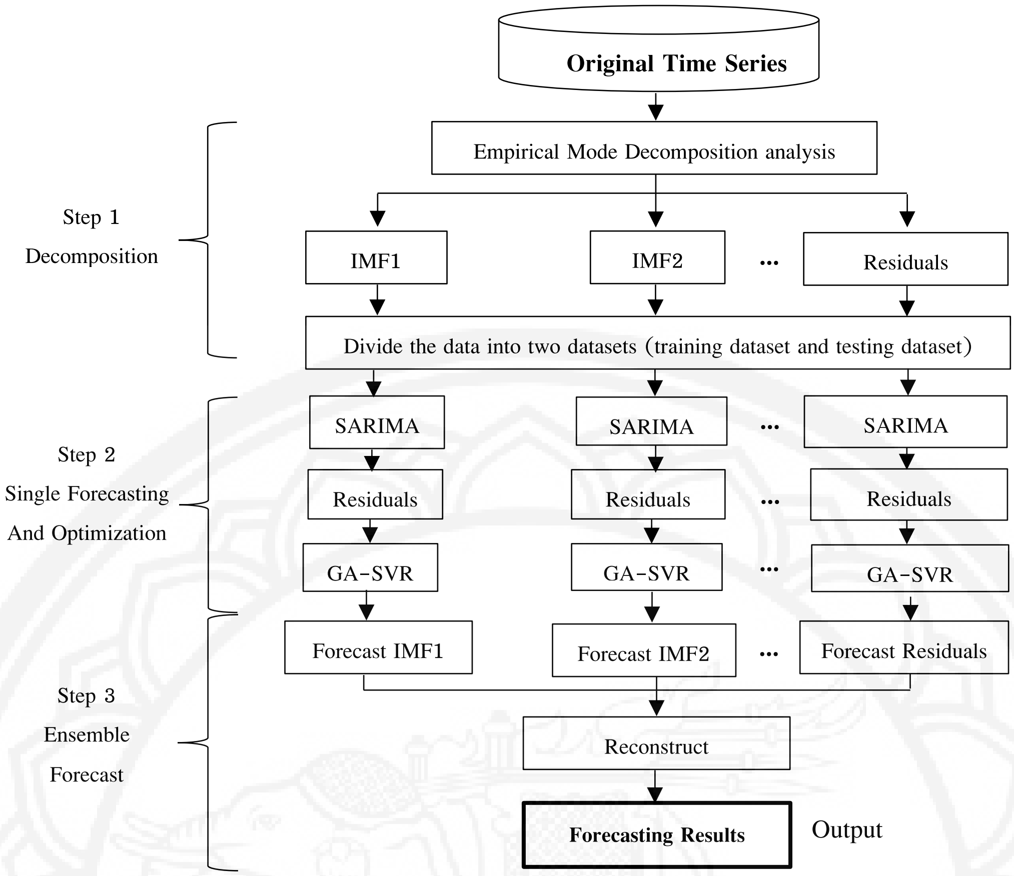

Figure 4 The Flowchart of the proposed hybrid forecasting system

Results

1. Data analysis results and basic statistics

The average monthly export volume of jasmine rice was nearly 156,000 tonnes with a standard deviation of 46,263 tonnes. The highest export volume was 336,287 tonnes while the lowest export volume was 67,000 tonnes. The average monthly export volume of Pathum Thani rice was 24,100 tonnes and the highest export volume was 92,643 tonnes while the lowest export volume was 3,238 tonnes. The average monthly export volume of sticky rice was 29,032 tonnes with the highest export volume of 112,791 tonnes while the lowest export volume was 5,458 tonnes, as presented in Table 1 (in kgs).

Table 1 Descriptive summary of three datasets of Thailand rice exports

Data set	data	Number	Mean	Std.	Max.	Min.	Median
Jasmine rice	All sample	135	155,644,571	46,263,254	336,287,581	66,999,988	149,362,875
	Training	108	164,420,544	43,015,954	336,287,581	94,499,283	154,523,230
	Testing	27	120,540,677	120,540,677	258,171,543	66,999,988	115,421,782
Pathum Thani rice	All sample	135	24,100,100	19,453,575	92,643,362	3,237,788	17,426,207
	Training	108	19,627,715	16,088,573	69,002,316	3,237,788	14,202,282
	Testing	27	41,989,637	21,674,035	92,643,362	3,877,297	38,970,697
Sticky rice	All sample	135	29,032,243	15,917,441	112,790,516	5,458,431	26,595,443
	Training	108	30,298,285	16,708,481	112,790,516	5,458,431	27,578,702
	Testing	27	23,968,074	11,117,908	52,985,167	9,312,299	22,371,168

2. Data analysis results of the SARIMA model

2.1 When considering the 3 sets of Thailand's rice export volume data (Figure 1), the graph tended to be a component with uneven distribution, indicating that the data were inconsistent. The results of the data consistency test with Augmented Dickey-Fuller (ADF) (Table 2) revealed that the ADF values of the 3 datasets were not statistically significant (p-value > 0.05), indicating that all 3 datasets had trends or seasonality components. When these three time series datasets were shown to be inconsistent, the data were transformed by identifying the logarithm of, and the differences between, the datasets before creating the SARIMA model. The results of the ADF analysis (Table 2) revealed that after transforming all 3 sets of data, they were consistent, and the ADF values were statistically significant (p-value < 0.05). Therefore, the data can be used to define the model.

Table 2 Augmented Dickey - Fuller (ADF) of the three datasets

data	Pre data transformation		Post data transformation	
	ADF	p-value	ADF	p-value
Jasmine rice	-3.1875	0.0936	-6.1166	0.01
Pathum Thani rice	-2.7276	0.2748	-5.3941	0.01
Sticky rice	-2.3531	0.4302	-6.9574	0.01

2.2 To define the pattern of the SARIMA model, (p,d,q) and (P,D,Q)s must be specified. The R programing language with an auto.arima( ) function in the Package "forecast" for determining the appropriate (p,d,q) and (P,D,Q)s was used to define the pattern of the SARIMA model. The results of the data analysis are presented in Table 3.

Table 3 Estimation of SARIMA of the three datasets

data		Estimate	Std. Error	z-value	p-value
Jasmine rice SARIMA(1,1,1)$(0,0,2)_{12}$	AR(1)	0.2453	0.1225	2.0022	0.045267
	MA(1)	-0.8475	0.0656	-12.9152	< 0.001
	SMA(1)	0.2136	0.0962	2.2197	0.026440
	SMA(2)	0.2854	0.1014	2.8150	0.004877
Pathum Thani rice SARIMA(0,1,1)$(0,0,1)_{12}$	MA(1)	-0.7017	0.0705	-9.9581	< 0.001
	SMA(1)	0.3627	0.1079	3.3606	< 0.001
Sticky rice SARIMA(1,1,0)$(0,0,1)_{12}$	AR(1)	0.37139	0.088970	4.1743	< 0.001
	SMA(1)	0.1826	0.093159	1.9601	0.04998

2.3 From Table 4 and Figure 4, the examination of the preliminary agreement of the SARIMA model revealed that the mean error value was zero. The errors were normally distributed, independent, and had constant variance. It can be concluded that the SARIMA model is suitable for forecasting the 3 datasets of Thailand's rice export.

Table 4 Examining the basic assumption of the SARIMA model of the three datasets

Models	The mean error value is 0		The error value is a normal distribution		The errors are dependent Ljung – Box Q-statistics		
	t-value	p-value	Shapiro-Wilk test	p-value	Q-statistics	df	p-value
Jasmine rice SARIMA(1,1,1)(0,0,2)$_{12}$	0.65228	0.5156	0.99455	0.9491	12.299	19	0.8725
Pathum Thani rice SARIMA(0,1,1)(0,0,1)$_{12}$	0.43801	0.6623	0.98212	0.1559	14.413	19	0.7591
Sticky rice SARIMA(1,1,0)(0,0,1)$_{12}$	-0.60027	0.5496	0.98656	0.354	29.308	19	0.0613

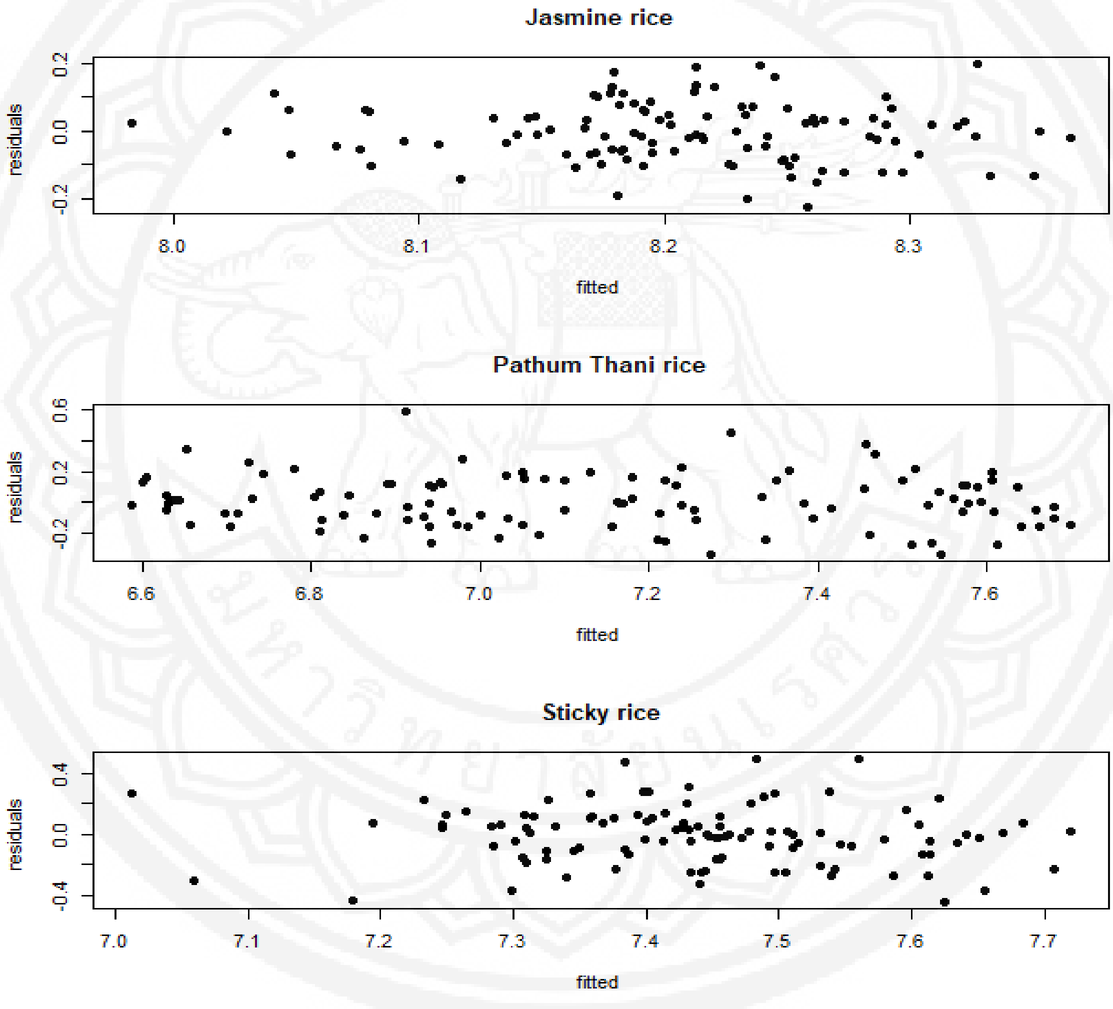

Figure 5 Plot residuals and fitted value of the three datasets

3. Results of the hybrid model

The researchers followed Step 6 in their quest to construct a sophisticated hybrid model. In the initial stage (Step 1), a meticulous analysis of the data was carried out using the EMD technique. This method effectively dissected the original time series into Intrinsic Mode Functions (IMF) and residuals, represented in Figure 6, as an example of EMD data analysis. Subsequently, the researchers carefully fashioned the hybrid model by making predictions for both the

Intrinsic Mode Functions (IMF) and the residuals, as showcased in Figures 3 and 4. The repertoire of hybrid models presented in this research comprises the distinguished models GA-SVR, EMD-SARIMA, EMD-GA-SVR, and EMD-SARIMA-GA-SVR. Finally, the selection of the most fitting model for forecasting Thailand's rice exports was accomplished through comparisons of their performance, as illustrated in Tables 5-6, along with visual representations displayed in Figures 7-9.

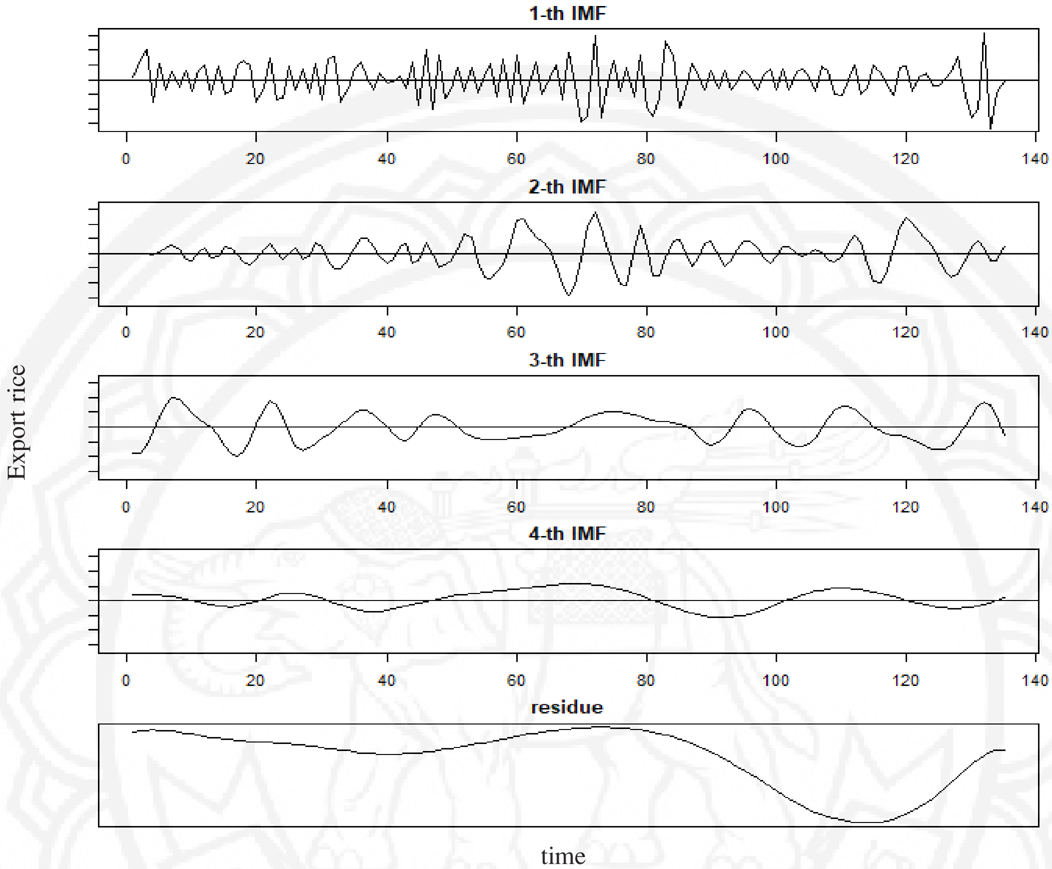

Figure 6 EMD components of export Jasmine rice dataset

Table 5 Comparison of the performance of the forecasting model of the three datasets

Data set	Models	*MAE*	*RMSE*	*MAPE*
Jasmine rice	SARIMA	29074934	38655587	23.94064
	GA-SVR	32970291	40916406	30.17244
	EMD-SARIMA	**20077642**	**25718196**	**16.67498**
	EMD-GA-SVR	21463307	26798637	18.01193
	EMD-SARIMA-GA-SVR	**19069569**	**25183275**	**15.49313**
Pathum Thani rice	SARIMA	15918746	21660669	97.06039
	GA-SVR	18210309	23973779	105.8153
	EMD-SARIMA	9637982	12157813	28.91725
	EMD-GA-SVR	10212778	12562637	44.70585
	EMD-SARIMA-GA-SVR	**8284756**	**10529064**	**21.26478**
Sticky rice	SARIMA	7750526	9671567	43.31443
	GA-SVR	7185445	10174671	28.47212
	EMD-SARIMA	8343842	10765820	47.59635
	EMD-GA-SVR	7392057	9358540	39.69793
	EMD-SARIMA-GA-SVR	**5127419**	**6878225**	**23.44107**

Table 6 Percentage improvement of the EMD-SARIMA-GA-SVR model in comparison with those of other forecasting models

Data set	Models	*MAE*	*RMSE*	*MAPE*
Jasmine rice	SARIMA	34.41233951	34.85217286	35.28523047
	GA-SVR	42.16135672	38.45188896	48.65138517
	EMD-SARIMA	5.020873467	2.079932045	7.087564723
	EMD-GA-SVR	11.15269888	6.027776711	13.98406501
Pathum Thani rice	SARIMA	47.95597593	51.39086424	78.09118632
	GA-SVR	54.50513223	56.08091657	79.90387023
	EMD-SARIMA	14.0405533	13.39672686	26.46333936
	EMD-GA-SVR	18.87852649	16.18746924	52.43401031
Sticky rice	SARIMA	33.84424489	28.8820002	45.88161497
	GA-SVR	28.64159422	32.39855127	17.67009271
	EMD-SARIMA	38.54846484	36.11053315	50.75027812
	EMD-GA-SVR	30.63610034	26.50322593	40.95140477

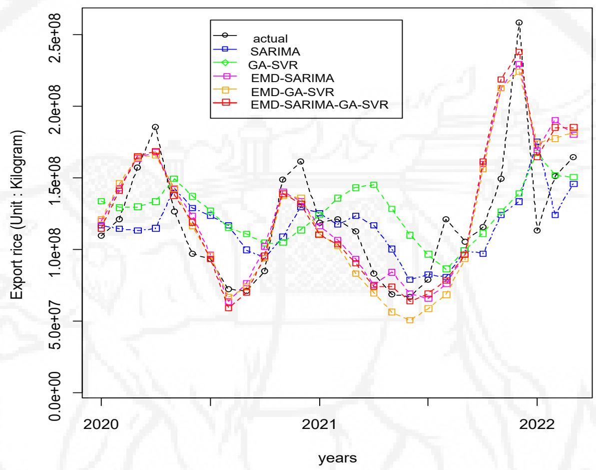

Figure 7 Original data and forecast results of jasmine rice from January 2020 to March 2022

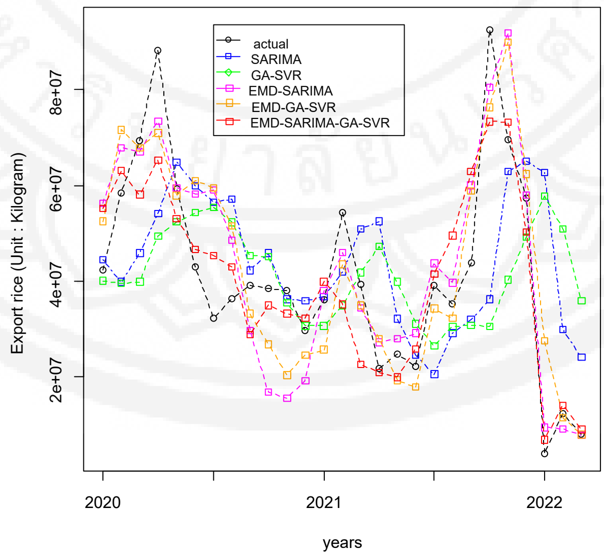

Figure 8 Original data and forecast results of Pathum Thani rice from January 2020 to March 2022

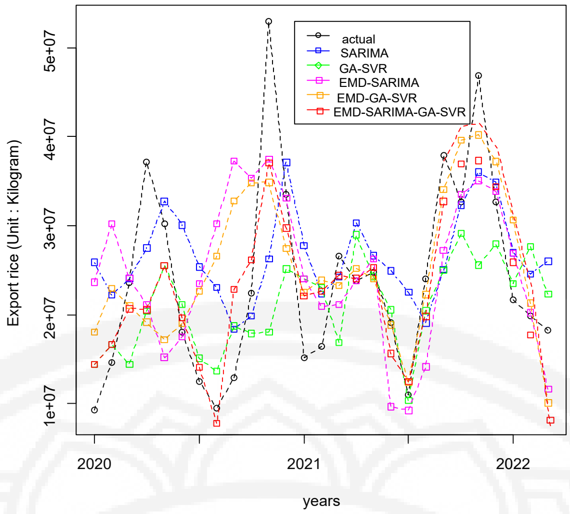

Figure 9 Original data and forecast results of sticky rice from January 2020 to March 2022

Discussion

In light of the test results obtained from the advanced forecasting model designed for predicting the quantity of rice exports in Thailand, a comprehensive analysis was conducted using separate datasets for jasmine rice, Pathum Thani rice, and sticky rice. The model itself was meticulously constructed through the integration of the Empirical Mode Decomposition (EMD) technique with the SARIMA and SVR models, representing a novel approach in this field of study. Following the model's development, a comparative analysis of its performance and accuracy was undertaken against a univariate time series model. This approach allowed for a comprehensive assessment of the forecasting model's efficacy and precision. The findings are discussed as follows:

1. The data utilized in this research comprises three datasets: jasmine rice, Pathum Thani rice, and sticky rice. All three datasets represent univariate time series data with monthly intervals. Plotting these datasets on a time series line graph, as illustrated in Figure 1, revealed that they exhibit similar movement patterns. Specifically, they demonstrate recurring seasonal patterns when tested for stationarity. However, upon thorough examination, it was determined that all three datasets are non-stationary time series. This implies that their mean and variance vary throughout the observed period. To address this issue, differencing the data is necessary to create a new time series before constructing the forecasting model. The test results for stationarity align with the appropriate analysis of the SARIMA model for the three time series patterns, as evidenced by Table 2 and Table 3.

2. The SARIMA and GA-SVR models: As shown in Table 5, the analysis of the time series data for rice exports from the three datasets using the SARIMA and GA-SVR models revealed that the SARIMA model outperformed the GA-SVR model for both the jasmine rice and Pathum Thani rice datasets. This finding aligns with the comparison graph depicting the forecasts made by the SARIMA model and the actual values within the same time intervals for the two datasets (as shown in Figures 7 and 8). It suggests that the SARIMA model effectively captures the movement patterns of both datasets, indicating a more pronounced linear relationship compared to a non- linear relationship. Consequently, the SARIMA model demonstrates superior performance in linear time series forecasting when compared to the GA- SVR model. However, the time series data for sticky rice exports showed that the GA-SVR model outperformed the SARIMA model, consistent with the comparison graph of forecasts made by the GA- SVR model and the actual values (Figure 9). This indicates that the time series data for sticky rice exports

exhibit more prominent non-linear relationships rather than linear relationships. Therefore, the GA-SVR model, renowned for its proficiency in non-linear time series forecasting, offers a more accurate explanation of the data characteristics compared to the SARIMA model. These findings corroborate the accuracy of the SARIMA model in linear time series forecasting (Basir et al., 2021) and the GA-SVR model in non-linear time series forecasting (Chen et al., 2021; Fan et al., 2021; Feng et al., 2020), as reported in the literature.

3. EMD-SARIMA, EMD-GA-SVR, and EMD-SARIMA-GA-SVR models: Based on the results of comparing the forecasting values of all three models for predicting the volume of rice exports from Thailand, as presented in Table 5 and Figure 7-9, it was determined that the EMD-SARIMA-GA-SVR model yielded the lowest error among the three sets. This indicates that utilizing the EMD method for data analysis reduces the volatility of time series data. Next, the decomposed components were forecasted using the SARIMA model, known for its strengths and efficiency in linear time series forecasting. The residual component from the SARIMA model was then further forecasted using the SVR model, which performs well in non-linear time series forecasting. Finally, the forecasts from both models were combined in the last step, resulting in highly accurate predictions and effectively reducing forecast errors in predicting the quantity of rice exports from Thailand. Considering the improvement percentages in Table 6 (specifically MAPE values), it was observed that the EMD-SARIMA-GA-SVR model achieved a greater reduction in forecast errors compared to the EMD-SARIMA and EMD-GA-SVR models for all three datasets. For Dataset 1 (Hom Mali rice), the percentage change values were 7.087564723 and 13.98406501. For Dataset 2 (Pathum Thani rice), the percentage change values were 26.46333936 and 52.43401031. As for Dataset 3, the percentage change values were 50.75027812 and 40.95140477. These findings align with the research conducted by Kao et al. (2020) on Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition, which concluded that the EEMD-ARIMA-GA-SVR model exhibited higher accuracy than other hybrid models. Similarly, Fan et al. (2022) investigated the Applications of Hybrid EMD with PSO and GA for an SVR-Based Load Forecasting Model and found that the EMD-PSO-GA-SVR model outperformed the SVR-PSO and SVR-GA models.

When comparing the EMD-SARIMA and EMD-GA-SVR models (Table 5), it was observed that the EMD-SARIMA model demonstrated higher accuracy than the EMD-GA-SVR model for Dataset 1 (Jasmine rice) and Dataset 2 (Pathum Thani rice) with MAPE values of 16.67498 and 28.91725, respectively. These findings suggest that these two datasets exhibit a more pronounced linear relationship, and the EMD-SARIMA model, known for its proficiency in linear forecasting, provides highly accurate predictions. However, for Dataset 3 (Sticky rice), the EMD-GA-SVR model outperformed the EMD-SARIMA model, indicating a stronger non-linear relationship in this dataset that favors the accuracy of the EMD-GA-SVR model (MAPE = 39.69793).

Nevertheless, despite the EMD-SARIMA-GA-SVR model developed in this research exhibiting high accuracy when compared to all the studied forecasting models, it possesses limitations in terms of computational time and model complexity. It requires more processing time and involves additional steps in model development to achieve lower errors compared to the original models. Furthermore, an intriguing observation is that when applying the EMD method to reduce data volatility before constructing the forecasting model, forecasting time series data may result in unusually high values of MAE, RMSE, and MAPE. This can be attributed to the forecasting of multiple intrinsic mode functions (IMFs), which contributes to cumulative forecast errors.

Conclusion and Suggestions

The results of from the study showed that the EMD-SARIMA-GA-SVR hybrid model outperforms the SARIMA, GA-SVR, EMD-SARIMA, and EMD-GA-SVR models in all performance metrics. In essence, the proposed EMD-SARIMA-GA-SVR hybrid model proved to be highly accurate in forecasting Thailand's rice export data. However, it should be noted that this hybrid model is more complex, involving multiple computational steps and longer processing time compared to the other models mentioned. Despite this limitation, if accuracy is a priority for organizations involved in rice export forecasting in Thailand, the hybrid model presents an intriguing option. Researchers and interested individuals can further develop or explore alternative hybrid models by incorporating EMD with ANN models or applying it alongside Variational Mode Decomposition (VMD) for SVR models, along with the use of the GA method. These approaches have the potential to enhance the accuracy of the forecasts. It is worth noting that when aiming for high accuracy, researchers must exercise caution in selecting input variables for model learning, as these factors greatly influence the forecasting model's precision.

Acknowledgements

We would like to express our sincere gratitude to the Faculty of Science, Ubon Ratchathani Rajabhat University for supporting the research with funding and the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives for providing the data used in this experiment. We also wish to thank Mr Roy I. Morien of the Naresuan University Graduate School for his editing of the grammar, syntax and general English expression in this manuscript.

References

Basir, M. S., Chowdhury, M., Islam, M. N., & Ashik-E-Rabbani, M. (2021). Artificial neural network model in predicting yield of mechanically transplanted rice from transplanting parameters in Bangladesh. *Journal of Agriculture and Food Research*, *5*, 1-8. https://doi.org/10.1016/j.jafr.2021.100186

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2008). *Time Series Analysis: Forecasting and Control* (4th ed.). New Jersey: John Wiley & Sons.

Chen, W., Xu, H., Chen, Z., & Jiang, M. (2021). A novel method for time series prediction based on error decomposition and nonlinear combination of forecasters. *Neurocomputing*, *426*, 85-103. https://doi.org/10.1016/j.neucom.2020.10.048

Co, H. C., & Boosarawongse, R. (2007). Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers & Industrial Engineering*, *53*, 610-627. https://doi.org/10.1016/j.cie.2007.06.005

Duan, W. Y., Han, Y., Huang, L. M., Zhao, B. B., & Wang, M. H. (2016). A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Engineering*, *124*, 54-73. https://doi.org/10.1016/j.oceaneng.2016.05.049

Fan, G. F., Yu, M., Dong, S. Q., Yeh, Y. H., & Hong, W. C. (2021). Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy*, *73*, 1-18. https://doi.org/10.1016/j.jup.2021.101294

Feng, Z., Niu, W., Tang, Z., Jiang, Z., Xu, Y., Liu, Y., & Zhang, H. (2020). Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *Journal of Hydrology*, *583*, 1-12. https://doi.org/10.1016/j.jhydrol.2020.124627

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation. *Departmental Technical Report*, *2*, 1-6.

Gu, Y. H., Yoo, S. J., Park, C. J., Kim, Y. H., Park, S. K., Kim, J. S., & Lim, J. H. (2016). BLITE-SVR: New forecasting model for late blight on potato using support-vector regression. *Computers and Electronics in Agriculture*, *130*, 169-176. https://doi.org/10.1016/j.compag.2016.10.005

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology Control and Artificial Intelligence*. Ney York: University of Michigan Press.

Huang, N.E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H.H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis, *Royal Society*, *454*, 903-995. https://doi.org/10.1098/rspa.1998.0193

Huang, Y., Hasan, N., Deng, C., & Bao, Y. (2021). Multivariate Empirical Mode Decomposition based Hybrid Model for Day-ahead Peak Load Forecasting. *Energy*, *239*, 1-15. https://doi.org/10.1016/j.energy.2021.122245

Jeong, S., Ko, J., & Yeom, J. (2022). Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea. *Science of the Total Environment*, *802*, 1-12. https://doi.org/10.1016/j.scitotenv.2021.149726

Kao, Y. S., Nawata, K., & Huang, C. Y. (2020). Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition. *Mathematics*, *8*, 1-19. https://doi.org/10.3390/math8101722

Keerativibool, W. (2014). Forecasting Model for the Export Value of Thai Jasmine Rice. *Burapha Science Journal*, *19*, 78-90.

Liu, L.W., Ma, X., Wang, Y. M., Lu, C. T., & Lin, W. S. (2021). Using artificial intelligence algorithms to predict rice (Oryza sativa L.) growth rate for precision agriculture. *Computers and Electronics in Agriculture*, *187*, 1-8. https://doi.org/10.1016/j.compag.2021.106286

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2015). *Misc Functions of the Department of Statistics.* Retrieved from http://cran.r-project.org/web/packages/e1071/index.html

Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L., & Wang, L. (2019). A robust method for non-stationary streamflow prediction based on improved EMD-SVM model. *Journal of Hydrology*, *568*, 462-478. https://doi.org/10.1016/j.jhydrol.2018.11.015

Mi, X., Liu, H., & Li, Y. (2019). Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Conversion and Management*, *180*, 196–205. https://doi.org/10.1016/j.enconman.2018.11.006

Office of Agricultural Economics. (2022). *Export.* Retrieved from http://impexp.oae.go.th/service/export.php?S

Prado, F., Minutolo, M. C., & Kristjanpoller, W. (2020). Forecasting based on an ensemble Autoregressive Moving Average—Adaptive neuro—Fuzzy inference system – Neural network—Genetic Algorithm Framework. *Energy*, *197*, 1-13. https://doi.org/10.1016/j.energy.2020.117159

Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. J. (2017). Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, *54*, 246–255. https://doi.org/10.1016/j.asoc.2017.01.015

R Core Team. (2021). *forecast: Forecasting Functions for Time Serie s and Linear Models.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://cran.r-project.org/web/packages/forecast/index.html

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, *53*(4), 1-37. https://doi.org/10.18637/jss.v053.i04

Su, Y., Xu, H., & Yan, L. (2017). Support vector machine-based open crop model (SBOCM): Case of rice production in China. *Saudi Journal of Biological Sciences*, *24*, 537–547. https://doi.org/10.1016/j.sjbs.2017.01.024

Sujjaviriyasup, T. (2021). A hybridization of feedforward neural network and differential evolution to forecast fertilizer consumption emphasizing on selecting optimal architecture. *Songklanakarin Journal of Science and Technology*, *43*, 1160-1168.

Tang, L. H., Bai, Y. L., Yang, J., & Lu, Y. N. (2020). A hybrid prediction method based on empirical mode decomposition and multiple model fusion for chaotic time series. *Chaos, Solitons & Fractals*, *141*, 1-12. https://doi.org/10.1016/j.chaos.2020.110366

Vapnik, V. N. (1998). *The Nature of Statistical Learning Theory*. New York: Springer Verlag.

Willighagen, E., & Ballings, M. (2021). *R Based Genetic Algorithm. CRAN Repository.* Retrieved from https://cran.r-project.org/web/packages/genalg/

Xu, X., & Zhang, Y. (2021). Corn cash price forecasting with neural networks. *Computers and Electronics in Agriculture*, *184*, 1-13. https://doi.org/10.1016/j.compag.2021.106120

Yang, Y., Che, J., Deng, C., & Li, L. (2019). Sequential grid approach-based support vector regression for short-term electric load forecasting. *Applied Energy*, *238*, 1010–1021. https://doi.org/10.1016/j.apenergy.2019.01.127

Yaslan, Y., & Bican, B. (2017). Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. *Measurement*, *103*, 52–61. http://dx.doi.org/10.1016/j.measurement.2017.02.007

Zhang, Z., Ding, S., & Sun, Y. (2020). A support vector regression model hybridized with chaotic krill herd algorithm and empirical mode decomposition for regression task. *Neurocomputing*, *410*, 185–201. https://doi.org/10.1016/j.neucom.2020.05.075